Sampling from Archimedean Copulas

Niall Whelan†

Abstract. We develop sampling algorithms for multivariate Archimedean copulas. For exchangeable copulas, where there is only one generating function, we first analyse the distribution of the copula itself, deriving a number of integral representations and a generating function representation. One of the integral representations is related, by a form of convolution, to the distribution whose Laplace transform yields the copula generating function. In the infinite dimensional limit there is a direct connection between the distribution of the copula value and the inverse Laplace transform. Armed with these results, we present three sampling algorithms, all of which entail drawing from a one dimensional distribution and then scaling the result to create random deviates distributed according to the copula. We implement and compare the various methods. For more general cases, in which an N dimensional Archimedean copula is given by $N-1$ nested generating functions, we present algorithms in which each new variate is drawn conditional only on the value of the copula of the previously drawn variates. We also discuss the use of composite nested and exchangeable copulas for modelling random variates with a natural hierarchical structure, such as ratings and sectors for obligors in credit baskets.

† Trade Floor Risk Management, Bank of Nova Scotia, 40 King Street West, Toronto ON M5H 1H1, Canada

E-mail: `niall_whelan@scotiacapital.com`

Submitted to: *Quantitative Finance*

1. Introduction

Use of copulas for multidimensional distributions is a powerful method of analysing the dependency structure of random variables. The term "copula" was first introduced by Sklar [Sklar, 1959] although some of the ideas go back to Höffding [Höffding, 1940]. They are useful because they permit us to focus on the dependency structure of the distribution independently of the marginal distributions of the random variables. Recently they have been considered for basket credit derivatives in finance [Li, 2000, Embrechts et al., 2001b, Schönbucher and Schubert, 2001, Bouyé et al., 2000, Razak, 2003] and for portfolio credit risk [Frey and McNeil, 2003, and references therein], where we assume that the marginal distributions can be obtained from market conditions, such as bond spreads or credit default spreads, but where we need to impose some information on the mutual dependencies. By marrying the market observable marginal distributions with a model choice for the copula, we can construct the full multidimensional distribution. This is used for modelling both N'th to default baskets and collateralised debt obligations (CDO's) whose payoffs depend on the probabilities of joint default. However, modelling credit derivatives is just one application of copulas and while it originally motivated this paper, the conclusions of this paper are not limited to finance applications.

Archimedean copulas (see [Nelsen, 1999, Joe, 1997] for reviews) are one class which has attracted particular interest since they have a number of properties which make them simple to analyse. Commonly we want to randomly sample from the multivariate distribution in order to perform a numerical simulation and an algorithm for two dimensional Archimedean copulas is presented in [Embrechts et al., 2001a]. In the current paper, we present sampling algorithms for Archimedean copulas in arbitrary dimension. Similar results can also be found in [Frey and McNeil, 2003], the latter being similar in spirit to one of the algorithms we present in this paper. Our goal is to avoid, as much as possible, direct determination of high order derivatives of many variables. This consideration rules out three generic multivariate sampling methods. If one invokes direct sampling by drawing a random variate x_1, then drawing x_2 conditional on x_1 and so on, one ends up having to determine high order derivatives of multiple variables. Another generic method is the rejection algorithm which is based on finding a covering distribution [Devroye, 1986], but this again requires the multivariate probability distribution function (pdf) which is difficult to determine. Finally there is the Metropolis algorithm in which one chooses whether to accept consecutive random draws using an acceptance criterion based on the density of the multivariate distribution, the determination of which again requires high order derivatives of many variables. Furthermore, the last method can be very slow since it involves acceptance criteria which can result in exponentially many rejected draws for every accepted draw. Because of these difficulties, we seek algorithms which are specific to Archimedean copulas.

We begin with a brief review of copulas and for this we first need to introduce some notation. We will use the symbol P to refer to the pdf of its argument. If we have

N random variables $y_1 \cdots y_N$, the joint pdf is $P(\mathbf{y})$ (where we use the bold notation to denote all of the variables y_i considered together.) We use $I(\mathbf{y})$ to denote the cumulative distribution function (cdf) of $\mathbf{y}$ and $I^*(\mathbf{y})$ to denote the complement of $I(\mathbf{y})$:

$$\begin{aligned} I(\mathbf{y}) &= \int_{-\infty}^{y_1} \mathrm{d}y_1' \cdots \int_{-\infty}^{y_N} \mathrm{d}y_N' P(\mathbf{y}') \\ I^*(\mathbf{y}) &= \int_{y_1}^{\infty} \mathrm{d}y_1' \cdots \int_{y_N}^{\infty} \mathrm{d}y_N' P(\mathbf{y}') \end{aligned} \tag{1}$$

For one dimensional normalised distributions we have $I(y)+I^*(y) = 1$, which is a relation we will use often. $I(\mathbf{y})$ represents the probability that every component of the random variable is less than the corresponding value of y_i while $I^*(\mathbf{y})$ represents the probability that every component of the random variable is greater than the corresponding value of y_i. For each dimension we also introduce the marginal pdf $P(y_i)$ which can be obtained from $P(\mathbf{y})$ by integrating over all components but y_i. There is also the corresponding marginal cdf $I(y_i)$ which equals $I(\mathbf{y})$ with all components but y_i set to infinity (or the largest values of the support of the respective pdf's.) For notational simplicity, we will use the symbols P and I generically to denote distributions of different variables; the arguments of the functions will in almost all cases resolve any ambiguity. Where there is ambiguity, we introduce subscripts.

It is convenient to introduce new variables $x_i = I_i(y_i)$ which have a domain from zero to one. (We have introduced a subscript on the functions I_i for notational clarity.) These new variables have their own multivariate distribution as well as marginal pdf's and cdfs. We recall that under any monotonically increasing change of variables $w = f(z)$ with distribution functions $I_z(z)$ and $I_w(w)$, those distribution functions are related by

$$I_w(w) = I_z(z = f^{-1}(w)). \tag{2}$$

This follows from the fact that $\Pr(z < Z)$ equals $\Pr(w < W)$ where $W = f(Z)$ or equivalently $Z = f^{-1}(W)$. Applying this identity to our situation in which $x_i = I_i(y_i)$, we find

$$C_i(x_i) = I_i(y_i = I_i^{-1}(x_i)) = x_i, \tag{3}$$

(where to lessen notational confusion we reserve the symbol C_i to refer to the cdf of x_i.) This is a convenient relation since it means that the marginal pdf of x_i is uniform. In fact, this is often taken as the definition of a copula, a multivariate distribution defined on the unit cube with uniform marginal densities. By analogous logic, the joint cdf of $\mathbf{x}$ is

$$C(x_1, \cdots, x_N) = I\left(y_1 = I_1^{-1}(x_1), \cdots, y_N = I_N^{-1}(x_N)\right). \tag{4}$$

We use the symbol C to indicate that in addition to being a cdf it is also a copula. C has a number of very specific properties [Nelsen, 1999, Joe, 1997] and we will only ever express it as a function of the random variables $\mathbf{x}$ since under a change of variables it loses the property of being a copula. In addition to being the cdf of $\mathbf{x}$, C can also be understood as a random variable in its own right (since it is a function of the random variables $\mathbf{x}$.) Specifically it has its own pdf and cdf which will be derived below.

As mentioned, a common application of copulas is to assume a specific copula functional form and combine this with information about the marginal distributions of the variables y_i to construct a full multivariate distribution for the y_i. All information about the dependency amongst the variables is embedded within the choice of copula. This procedure can be thought of as extending information about the marginal distributions to knowledge about the complete distribution. It has the added benefit of cleanly separating the information about the marginal distributions from the assumptions about their dependence structure. For example, it is relatively simple to change the copula while keeping the marginal distributions constant and thereby explore the effect of different choices of dependence. Different copulas may also have properties not present for standard multivariate distributions such as the normal. For instance, the normal distribution has asymptotic independence unless correlations are unity. Other copulas may not be asymptotically independent, meaning that events in which two or more deviates are in the tails of their respective marginal distributions are relatively more common. This can be important when trying to model joint defaults, for example.

All copulas are bounded between the functions $M_N = \min(x_1, \cdots x_N)$ and $W_N = \max(x_1 + \cdots + x_N - N + 1)$, known as the Fréchet-Höffding bounds. M_N is the maximum and is itself a copula (corresponding to perfect dependence amongst the x_i) while W_N is the minimum but is only a copula for $N = 2$. Standard copulas include the product copula $\Pi = \prod_i x_i$ which corresponds to statistical independence. There is also the Gaussian copula which depends parametrically on a correlation matrix. Assuming the marginal distributions are all Gaussians, this corresponds to a multivariate Gaussian distribution. Another class are extreme value copulas which have attracted interest in finance applications [Bouyé et al., 2000]; in particular they lead to first to default intensities which are free of term structure [Razak, 2003]. There are also the Archimedean copulas which have a number of attractive properties and which will be explored in the sections to come.

In the following section we analyse various features of the distribution of copula values for one simple but commonly considered class of Archimedean copulas, which are called "exchangeable". In the section following that we use these results to provide a number of sampling algorithms. In the two subsequent sections we discuss two classes of copulas which generalise the exchangeable copulas in different ways and show how to sample from them. In the second to last section we introduce a class of inexchangeable Archimedean copulas which have properties of both of the generalisations mentioned above and which can reflect hierarchical structures which may exist among the random variables. An example is a set of obligors organised by credit rating, industrial sector and home country. We end with a conclusion.

2. Analysis of Exchangeable Archimedean Copulas

As mentioned in the introduction, one commonly considered class are the Archimedean copulas [Nelsen, 1999, Joe, 1997] for which the copula function has the form

$$C(x_1, x_2, \cdots x_N) = \phi^{-1}\left(\phi(x_1) + \phi(x_2) + \cdots + \phi(x_N)\right), \tag{5}$$

where $\phi(x)$ is a generating function which is defined in the range zero to one, is monotonically decreasing, and $\phi(1) = 0$. Furthermore, if $\phi(0) = \infty$ the copula is said to be a "strict Archimedean". Most useful copulas are strict and we assume this condition in what follows. Examples include [Marshall and Olkin, 1988] the independent copula for which $\phi(x) = -\log(x)$, the Gumbel copula for which $\phi(x) = (-\log(x))^\alpha$ with $1 \leq \alpha < \infty$, and the Clayton copula for which $\phi(x) = (x^{-\alpha} - 1)/\alpha$ with $0 < \alpha < \infty$. Note that these are all strict. The Gumbel copula is privileged in being both an Archimedean and an extreme value copula. There is a permutation symmetry among the x_i in (5) and the term "exchangeable" is used describe this situation and to distinguish it from more general classes of Archimedean copulas which will be presented below.

Because it arises so often in the subsequent discussion it is useful to define the function $f(x) = \phi^{-1}(x)$. (Note that some authors define ϕ and f oppositely to what we have done here; while arguably a superior convention we have elected to follow the convention in the finance literature.) Archimedean copulas arise naturally in the context of Laplace transforms of distribution functions [Joe, 1997]. If $f(x)$ is the Laplace transform,

$$f(x) = \int_0^\infty \mathrm{d}s\, e^{-sx} F(s) \tag{6}$$

of some univariate density function $F(s)$, then the function (5) is guaranteed to be a proper distribution, meaning that its density function as well as all marginal density functions are positive. If, on the other hand, we do not obtain ϕ from a Laplace transform, then in general we are not guaranteed that (5) is a proper distribution. A necessary condition is that $f(x)$ have negative first derivative everywhere and that subsequent derivatives alternate in sign [Kimberling, 1974]. This property is referred to as being "completely monotone" [Schoenberg, 1938] and implies that it is analytic in the right half complex plane $\{z = a + ib, a > 0\}$ and typically, although not necessarily, has a singularity at the origin. This condition is well known and can be seen to arise quite naturally in the subsequent discussion. The need for this additional condition arises because we are specifying the cdf (*i.e.* the copula). Traditional routes to copulas start with some pdf (*e.g.* the multivariate normal distribution) and then derive from it the corresponding copula. One is then left with a copula whose density is necessarily positive but which is algebraically intractable. With Archimedean copulas on the other hand, the copula is quite straight-forward to analyse but unless $f(x)$ is defined from a Laplace transform as in (6) one must be manifestly concerned with guaranteeing positivity of the densities.

2.1. The Distribution Function of the Copula

In this subsection we develop expressions for the distribution of the copula value. It is actually simplest to work in terms of the variable $t = \phi(C)$, as we now show. We start by defining a new set of variables ξ where

$$\xi_i = \phi(x_i), \tag{7}$$

in terms of which $C(\mathbf{x}) = f\left(\sum_i \xi_i\right)$. Recall that $C(\mathbf{x})$ can be interpreted as the probability that each component of $\mathbf{x}$ is smaller than the corresponding argument of C. Due to the fact that ξ_i is a decreasing function of x_i, this equals the probability that each component of ξ is greater than the corresponding argument. Symbolically, $I^*(\xi) = C(\mathbf{x}) = f\left(\sum_i \xi_i\right)$. We can then differentiate to find the pdf of ξ.

$$P(\xi) = (-1)^N \frac{\partial^N}{\prod_i \partial \xi_i} f\left(\sum_i \xi_i\right). \tag{8}$$

The factor $(-1)^N$ arises from the definition of I^* in terms of P, there is a negative sign per component.

We now introduce a new change of variables, analogous to hyper-spherical polar coordinates: $t = \sum_i \xi_i$ and a set of $N-1$ variables u_i which span the constant t surface. (These are known as "Jacobi coordinates" in the context of dynamical systems in which we use centre of mass coordinates.) Since P only depends on t, we can focus most of our attention on it and we actually do not need to be very specific about the parameterisation of the u_i. In particular,

$$C(\mathbf{x}) = f(t). \tag{9}$$

As mentioned, most of our results are for the distribution of t, but this is trivially mapped to the distribution of the copula value C.

The pdf of the set $(t, \mathbf{u})$ is related to that of the ξ_i in (8) by the Jacobian of the transformation. We can always define the u_i such that this Jacobian is unity so that

$$P(t, \mathbf{u}) = (-1)^N f^{(N)}(t) \tag{10}$$

where the superscript on f indicates the number of derivatives to take. This is the joint pdf of both t and $\mathbf{u}$. Although there is no explicit dependence on the latter, they still play a role when we integrate over their measure. More specifically, if we want the marginal pdf of just t, we need to integrate over all the $\mathbf{u}$:

$$\begin{aligned} P_N(t) &= \int \prod_i \mathrm{d}u_i P(t, \mathbf{u}) \\ &= (-1)^N \frac{t^{N-1}}{(N-1)!} f^{(N)}(t) \end{aligned} \tag{11}$$

where the factor $t^{N-1}/(N-1)!$ comes from the integral over the u_i. That the $\mathbf{u}$ integral leads to this factor is trivially true for $N = 1$ (for which there is no integral) and larger values of N follow by induction; the proof is trivial and we do not present it. Also, we have introduced a subscript of N on P to specify for which dimension it is relevant.

Equation (11) has an interpretation as a generalisation of a Poisson distribution. In fact, for the independent copula for which $f(x) = \exp(-x)$, the formula is exactly the pdf of the N'th arrival time of a Poisson process.

A further comment on (11) is on the alternating signs as a function of N. In order for the left hand side to be positive, we clearly need that alternate derivatives of the function $f(t)$ alternate in sign. As mentioned earlier, this is the property of being completely monotone and is a condition for the function ϕ to generate a proper distribution.

One nice property of $P_N(t)$ is that it is a perfect derivative, although this is not necessarily obvious. Nevertheless it can be expressed as

$$P_N(t) = -\frac{\mathrm{d}}{\mathrm{d}t} I_N^*(t). \tag{12}$$

where

$$I_N^*(t) = \sum_{n=0}^{N-1} \frac{(-t)^n}{n!} f^{(n)}(t). \tag{13}$$

This is easy to see by differentiating the series (13) to find a telescoping series of which (11) is the only surviving term. Recall that $I_N^*(t)$ is the integral from t to ∞ of the pdf of t and is the complement of the normal cdf. In particular, $I_N^*(\infty) = 0$ while $I_N^*(0) = f(0) = 1$. The latter in particular indicates that the pdf is correctly normalised‡. The expressions (11) and (13) are the main results of this section and are what we shall build on in what follows.

We can also derive a generating function expansion for the $P_N(t)$ by defining§

$$g(t,z) = \sum_{N=1}^{\infty} z^{N-1} P_N(t). \tag{14}$$

We use the expression (11) and identify the sum as the Taylor expansion of $f'(x)$ expanded around $x = t$ and evaluated at $x = (1-z)t$ to conclude

$$g(t,z) = -f'((1-z)t), \tag{15}$$

in terms of which we can find any $P_N(t)$ as

$$P_N(t) = \frac{1}{(N-1)!} \frac{d^{N-1}}{dz^{N-1}} g(t,z)\bigg|_{z=0}. \tag{16}$$

We can integrate this with respect to t to find

$$I_N^*(t) = \frac{1}{(N-1)!} \frac{d^{N-1}}{dz^{N-1}} h(t,z)\bigg|_{z=0}, \tag{17}$$

‡ It is interesting to note that (13) looks much like a Taylor series expansion. In fact if we introduce the complementary series $I_N(t) = \sum_{n=N}^{\infty} \frac{(-t)^n}{n!} f^{(n)}(t)$ then $I_N(t) + I_N^*(t)$ together represent $f(0)$ Taylor expanded around the point t. The series can be shown to converge using the Laplace relations (6). However, by definition $f(0) = 1$ and so $I_N(t) = 1 - I_N^*(t)$ and is therefore an equally valid representation of the cdf of t. This is reminiscent of the phenomenon of "resurgence" in asymptotic analysis where there is a complementarity between the so-called "early terms" and "late terms" in a properly resummed asymptotic expansion.

§ I wish to thank Tom Hurd for this idea.

where $h(t,z) = f((1-z)t)/(1-z)$. The functions $g(t,z)$ and $h(t,z)$ have analyticity properties governed by that of the function $f(z)$. In particular, there is generally a singularity at $t = 0$ and at $z = 1$. Although the expression (14) is typically only convergent for $|z| < 1$, (15) is the unique analytic continuation and is valid for all values of z, although care must be taken since the functions g and h may be multi-valued in the complex plane.

The expression (13) is directly related to an expression due to Barbe et al. [Barbe et al., 1996] for the cdf of the copula itself. We can think of the expression (9) as defining a change of variables from t to C which can be made explicit by writing $t = \phi(C)$. Since $f(t)$ is a decreasing function, the probability that t is greater than some value T is the same as the probability that C is less than $f(T)$. Therefore we can immediately write the cdf of C as

$$\begin{aligned} I_N(C) &= \sum_{n=0}^{N-1} \frac{(-1)^n \phi^n(C)}{n!} f^{(n)}\left(\phi(C)\right) \\ &= C + \sum_{n=1}^{N-1} \frac{(-1)^n \phi^n(C)}{n!} f^{(n)}\left(\phi(C)\right). \end{aligned} \tag{18}$$

This formula can be found in [Barbe et al., 1996] where is is refered to as $K(t)$. The exposition above is an alternate and arguably simpler route to the same result. In two dimensions this simplifies to

$$I_2(C) = C - \frac{\phi(C)}{\phi'(C)}. \tag{19}$$

We conclude that the expressions (11) and (13) can also be understood as determining the cdf of the copula itself. This fact is potentially useful in fitting empirical data to an Archimedean copula. In [Genest and Rivest, 1993] and [Bouyé et al., 2000] the authors discuss parametric fitting of empirical data. In particular, one can estimate the distribution function of the copula using nonparametric estimation and then do a parametric estimation of the best-fitting Archimedean copula. The distribution of t might potentially aid in this endeavour as well as being useful for sampling. However to proceed with either of these applications, it is important to have a computationally tractable representation of the distribution. Most generating functions are sufficiently complicated that the higher derivatives get very complicated quickly and it is not feasible to determine them directly. Therefore in the next two subsections we present two integral representations of the distribution which are simpler to work with than (13).

2.2. Cauchy Integral Representation

To obtain the first integral representation, we make use of the Cauchy formula to determine the derivatives in (13). Recall that if we have a function $f(t)$ we can express any derivative as

$$f^{(n)}(t) = \frac{n!}{2\pi i} \oint \mathrm{d}z \frac{f(z)}{(z-t)^{n+1}} \tag{20}$$

where the integration contour encloses the pole at $z = t$ once and otherwise crosses no poles or branch-cuts of f. We can insert this expression into (13) and sum a geometric series which is generated to conclude

$$I_N^*(t) = -\frac{1}{2\pi i}\oint \mathrm{d}z \frac{f(z)}{z}\left(\frac{t}{t-z}\right)^N. \tag{21}$$

This is a usable representation of the cdf since we can readily perform the contour integral numerically. This is discussed in the appendix. As mentioned, $f(z)$ being completely monotone means that it is only guaranteed to be analytic in the positive real part of the complex plane so the contour is limited to that domain. In particular, it should avoid the origin which typically has a singularity. Applying (20) to the equation in a previous footnote leads to the following alternate form

$$I_N(t) = 1 + \frac{1}{2\pi i}\oint \mathrm{d}z \frac{f(z)}{z}\left(\frac{t}{t-z}\right)^N. \tag{22}$$

The integral representations above are particularly effective when t is large. When t is small they are not numerically efficient but this is overcome by changing variables from z to $1/z$ so that alternate representations of the same functions are

$$\begin{aligned} I_N^*(t) &= \frac{1}{2\pi i}\oint \mathrm{d}z \frac{f(1/z)}{z}\left(\frac{z}{z-1/t}\right)^N \\ I_N(t) &= 1 - \frac{1}{2\pi i}\oint \mathrm{d}z \frac{f(1/z)}{z}\left(\frac{z}{z-1/t}\right)^N \end{aligned} \tag{23}$$

Again we must avoid the origin and limit the contour to the positive real half plane when encircling the pole at $z = 1/t$.

2.3. Laplace Integral Representation and $N \to \infty$

Another path to an integral representation is to appeal to the definition of $f(t)$ as a Laplace transform. In particular we will see that this leads to an interesting relation in the limit of large dimension. Combining (6) and (11) we find

$$P_N(t) = \frac{1}{\Gamma(N)}\int_0^\infty \mathrm{d}s\, s e^{-st}(st)^{N-1}F(s). \tag{24}$$

We can integrate this with respect to t to find

$$\begin{aligned} I_N(t) &= \int_0^\infty \mathrm{d}s F(s)\gamma(N, st) \\ I_N^*(t) &= \frac{t}{\Gamma(N)}\int_0^\infty \mathrm{d}s e^{-st}(st)^{N-1}I(s) \end{aligned} \tag{25}$$

where $\gamma(N, x)$ is the incomplete gamma function and is itself a cdf while $I(s)$ is the cdf corresponding to $F(s)$. One can see by inspection that $I_N(\infty) = 1$, assuming $F(s)$ is normalised. This is a central relation which we will refer to often in the subsequent discussion. We can think of this as a form of convolution in which the cdf $\gamma(N, x)$ is

spread out by integration against a broadening function‖. In fact, we can immediately understand the infinite dimensional limit. We earlier remarked that (11) resembles the distribution of the N'th arrival time for some process. We know what happens in the infinite dimensional limit for the Poisson generating function (*i.e.* the independent copula) and it is natural to ask the same question for nonindependent copulas. To motivate this, it is useful first to look at the moments of the distribution of t.

Using (11) we find that the m'th moment of t is given by the following recursion relation

$$\begin{aligned}\langle t^m\rangle_N &= \frac{(-1)^N}{(N-1)!}\int_o^\infty \mathrm{d}t\, t^{m+N-1} f^{(N)}(t)\\ &= \frac{N+m-1}{N-1}\langle t^m\rangle_{N-1}.\end{aligned} \tag{26}$$

To obtain this, we have integrated by parts and identified the resulting integral as being proportional to the m'th moment for dimension $N-1$. We can obviously continue this and we find

$$\langle t^m\rangle_N = \left(\begin{array}{c} m+N-1 \\ N-1 \end{array}\right)\langle t^m\rangle \tag{27}$$

where we have defined $\langle t^m\rangle$ without a subscript with reference to $N=1$, that is

$$\langle t^m\rangle \equiv \langle t^m\rangle_{N=1} = -\int \mathrm{d}t\, f'(t) t^m. \tag{28}$$

In particular, the mean value of t is proportional to N so it is useful to scale this out by defining a new random variable $\tau = t/N$. The first two moments of τ are

$$\begin{aligned}\langle\tau\rangle_N &= \langle t\rangle\\ \langle\tau^2\rangle_N &= \frac{1}{2}\left(1+\frac{1}{N}\right)\langle t\rangle^2\end{aligned} \tag{29}$$

from which we can determine the variance

$$\sigma_\tau^2 = \left(\frac{1}{2}\langle t^2\rangle - \langle t\rangle^2\right) + \frac{1}{2N}\langle t\rangle^2. \tag{30}$$

There are two interesting limits. One is that the copula generating function is such that the first term in (30) is zero. In that case, only the second term is relevant and we find that the standard deviation decreases as $1/\sqrt{N}$ for large N. This case is already very familiar, being the distribution of the N'th arrival time of a Poisson process; this approaches a Gaussian centered on $\tau = 1$, which in the large N limit becomes a delta function. Therefore we will not dwell on this and rather focus on the other interesting limit suggested by (30) which is to take $N \to \infty$ so that the second term vanishes. However we would like to be more ambitious for this limit than just to examine the first few moments but rather derive an expression for the entire distribution. We first derive

‖ More precisely (25) is in the form of a correlation between the functions F and γ but where we multiply, rather than add, the arguments s and t in γ. In terms of $\log t$ and $\log s$, (25) is a standard correlation integral.

an expression for all moments by revisiting (27) but focusing on the large N limit and expressing the results in terms of τ so we can write

$$\begin{aligned}\langle \tau^m \rangle &= \frac{1}{N^m} \left(\begin{array}{c} m+N-1 \\ N-1 \end{array} \right) \langle t^m \rangle \\ &\approx \frac{1}{m!} \langle t^m \rangle. \end{aligned} \tag{31}$$

In the second line, the approximation is that $N \gg m$. The previous expression tells us that there is a deep connection between the infinite dimensional limit of the τ distribution and the one dimensional t distribution. Of course, the precise nature of that connection is embedded in the expressions (24) and (25).

In terms of τ, we can reexpress (25) as

$$I_N(\tau) = \int_0^\infty \mathrm{d}s F(s) \gamma(N, Ns\tau). \tag{32}$$

In the infinite dimensional limit, a saddle-point analysis indicates that

$$\lim_{N\to\infty} \gamma(N, Nx) = \Theta(x-1) \tag{33}$$

where $\Theta(x)$ is the Heaviside step function. (More precisely, for large N, the left hand side can be shown to approach a normal cdf which becomes infinitely sharp as $N \to \infty$.) Using this identity and defining $G(s)$ as the cdf of s, corresponding to the pdf $F(s)$, we conclude that

$$\begin{aligned} I^*_\infty(\tau) &= G(1/\tau) \\ P_\infty(\tau) &= \frac{1}{\tau^2} F(1/\tau), \end{aligned} \tag{34}$$

Equivalently, if we define $\omega = 1/\tau$, then the pdf of ω is simply $F(\omega)$. We have hereby proved that in the infinite dimensional limit, the distribution of τ is given by the distribution whose Laplace transform generates the Archimedean copula. As far as we are aware, this is a new result. The expressions (34) are the main result of this section and will prove of use in developing sampling algorithms.

We can also express this directly in terms of the copula itself through the change of variables $\tau = \phi(C)/N$ in the limit that $N \to \infty$ so that the distribution of C is

$$\begin{aligned} I_\infty(C) &= \lim_{N\to\infty} G\left(\frac{N}{\phi(C)}\right) \\ P_\infty(C) &= -\lim_{N\to\infty} N \frac{\phi'(C) F\left(\frac{N}{\phi(C)}\right)}{\phi^2(C)}, \end{aligned} \tag{35}$$

although these are probably less useful since they require care in understanding the limit.

We can explicitly verify (34) for the independent copula for which $f(t) = \exp(-t)$ and hence $F(s) = \delta(s-1)$. Using this relation in (34) we find $P_\infty(\tau) = \delta(\tau - 1)$, which is the correct limiting distribution of the average arrival time of a Poisson process in the limit that $N \to \infty$. The effect of using something other than the independent copula

is to broaden the distribution (as we observe from the variance being nonzero in (30).) The expression (34) is then useful for understanding the asymptotic behaviour of $P(\tau)$ from the characteristics of the Laplace transform, for example. We hope to explore this in a later publication as well as the case that N is large but not infinite in (25).

As a final piece of analysis, we can determine a relatively simple expression for the characteristic function of (34) using the moments of the distribution. In general, the characteristic function of a pdf $P(x)$ can be expressed in series form as

$$Q(k) = \sum_{m=0}^{\infty} \langle x^m \rangle \frac{(2\pi i k)^m}{m!}. \tag{36}$$

Applying this identity to our case together with the results in (28) and (31), we find

$$\begin{aligned} Q_\infty(k) = & -\int_0^\infty \mathrm{d}t f'(t) \sum_{m=0}^{\infty} \frac{(2\pi i k t)^m}{m!m!} \\ = & -\int_0^\infty \mathrm{d}t f'(t) J_0\left(e^{-i\pi/4}\sqrt{2\pi k t}\right), \end{aligned} \tag{37}$$

in terms of which

$$P_\infty(\tau) = \int_{-\infty}^{\infty} \mathrm{d}k\, e^{-2\pi i k \tau} Q_\infty(k). \tag{38}$$

The function $J_0(e^{-i\pi/4}z)$ can also be expressed in terms of Kelvin functions [Abramowitz and Stegun, 1965]. In fact, performing the above Fourier transform, leads to the expressions (34) and this constitutes an alternate route to the same result.¶

3. Sampling from Exchangeable Copulas

In this section we present two strategies for sampling from the copula using the results of the previous sections. One strategy is to find a random value of t and then partition it among the ξ (and hence among the $\mathbf{x}$.) The first two subsections present two algorithms for randomly sampling a value t while the third subsection presents an algorithm for partitioning a sampled value of t randomly among the $\mathbf{x}$. In the fourth subsection we present a slightly different strategy which makes use of the infinite dimensional limit of the τ distribution. In most cases this is probably the most efficient algorithm but to explain it, we find it helpful to start by explaining the first strategy. In all cases we make explicit use of the exchangeability property.

¶ Analysis of the Fourier integral requires some care since if we interchange the t and k integrals, the integral over k is pure imaginary and divergent. Rather, we break the k integral into positive and negative k domains and introduce a small imaginary component $\pm i\epsilon$ to τ to render the integrals convergent, the sign of ϵ depending on the sign of k. The two k integrals can then be found from tables [Oberhettinger, 1990, for example]. We then sum them and take the limit $\epsilon \to 0$ and the rest follows.

3.1. First Algorithm for Drawing t

The first algorithm is to integrate (25) in closed form. This only works for relatively simple copulas. One example is the Clayton copula for which [Marshall and Olkin, 1988]

$$F(s) = \frac{(1/\alpha)^{1/\alpha}}{\Gamma(1/\alpha)} e^{-s/\alpha} s^{1/\alpha - 1}. \tag{39}$$

Using this in (25) we find the complement of $I_N(t)$ as

$$I_N^*(t) = \frac{B\left(\frac{1}{1+\alpha t}; N, 1/\alpha\right)}{B\left(N, 1/\alpha\right)}. \tag{40}$$

where $B(z; a, b)$ is the incomplete Beta function and $B(a, b)$ is the complete Beta function. This is manifestly normalised and there are standard techniques to sample from it, as we discuss below.

3.2. Second Algorithm for Drawing t

For this algorithm we use the complex integral representations of subsection 2.2. We can for instance sample a random uniform deviate r and then find the value of t consistent with it. This technique is suitable for situations where we do not have numerically efficient means of directly determining $F(s)$. As discussed in the appendix, one virtue of having a closed contour representation is that the integral may be relatively inexpensive to calculate since it converges very rapidly with the number of points used.

We forego presenting the analogous method using the Laplace representation 25 since it explicitly requires $F(s)$. If we know $F(s)$, then there is a superior method which we present below. If we do not know $F(s)$ then using the Laplace representation would be difficult to implement.

3.3. Partitioning t among the ξ

For purposes of this subsection we imagine that using one of the previous algorithms, we have chosen a random value of t from its distribution. Given such a random value, we would like to randomly select a set ξ with the constraint that their sum equals t. We can think of this geometrically as picking points randomly on a hyperplane in N-dimensional space, also known as a simplex. Methods for sampling from it can be found in [Devroye, 1986]. One algorithm follows by analogy with sampling from a unit sphere. We make N uncorrelated draws from the simple Poisson pdf $\exp(-y)$, calling the draws r_i. We rescale each of them by the ratio of t and their sum. This works because the Poisson distribution has the property $\prod_i P(x_i) = P(\sum_i x_i)$, so that the independent multivariate density is a function of the sum of the random variables.

Symbolically, we find random deviates r_i drawn from the Poisson distribution and then define

$$\xi_i = \frac{r_i}{\sum_{i=1}^N r_i} t. \tag{41}$$

As a final step, we define $x_i = f(\xi_i)$ and we are done. As a result of this algorithm, we have a random draw of values from (5). For many copulas, most of the difficult numerical work is in finding a random value of t, whereas finding the N random values r_i is very fast even for large N.

3.4. Third algorithm using the distribution of τ

The expression (34) has an immediate application as the basis for an alternate sampling strategy. The idea is to conceptually imagine that we make an infinite dimensional draw but that we only record a number d of the variates. The set we record will perforce be distributed as the d dimensional copula.

We first make a random draw ω from the pdf $F(\omega)$ which we invert as $\tau = 1/\omega$, from which we can define $t = N\tau$. Recall that we have defined $t = \sum_{i=1}^{N} \xi_i$. In the previous subsection we described how to sample from the constant-t simplex in the space of ξ. It amounted to making one draw from Poisson distribution for each dimension, that draw being called r_i and then using (41). We now recall that $t = N\tau$ and also note that the denominator of (41) approaches N in the limit so we are left with

$$\xi_i = r_i \tau. \tag{42}$$

To recap, we first draw a value ω from $F(\omega)$ which we invert to get τ. For each dimension we draw a random variate r_i from the Poisson distribution and multiply it by τ. These will then be random deviates ξ_i. We then use $x_i = f(\xi_i)$ to determine the random deviates in the original defining space of the copula. In most cases this is probably more efficient than the previous algorithms since one is saved the necessity of using any integral representations. One may still need to precalculate $F(s)$ supplemented with interpolation if it is not known in closed form. However this is a relatively small cost and need only be born once, independently of the dimensionality or the number of samples required. For copulas where $F(s)$ is known, this method is very fast.

This algorithm is very similar in spirit to one proposed in [Frey and McNeil, 2003, Lemma 4.15]. The only difference is that we are working in terms of variables t and τ, which makes the sampling particularly simple, however the logic is the same. We also remark that this algorithm is quite similar to sampling from a multivariate Student's t distribution where one first draws from a χ^2 distribution, then takes the inverse of the square root of the draw and finally scales that number by draws from a normal distribution.

Another interesting observation is that we can also pursue the conceptual idea encapsulated in (42) to proceed with the formal development. That is we can define variables $\xi_i = r_i\tau$ and given the distributions of r_i and τ, we can determine the distribution of the ξ_i. From that we can then determine the distribution of the variables $x_i = f(\xi_i)$. We do not present the details but it is not difficult to show that one ends up with a Laplace transform of a product which in the end leads back to (5). In a sense

then we have come full circle but along the way have derived some useful alternate expressions for the distribution.

3.5. Numerical Results

For all algorithms we require sampling either from the univariate distribution of t or of τ (the latter being directly related to the distribution $F(s)$.) Two basic approaches are either to use the rejection algorithm [Devroye, 1986] or by direction inversion of the cdf. The rejection algorithm is the method of choice for many standard univariate distributions such as the gamma and beta distributions [Ahrens and Dieter, 1974]. In order to use this, one needs to find a covering distribution whose pdf exceeds the pdf of the desired distribution for all values. The covering distribution should also be simple to sample from. For our purposes we in fact need to find parametric families of covering distributions which work for all choices of parameter and of copula dimensionality (the latter constraint is only relevant for $P(t)$.) For example, for the Clayton copula we have $I_N(t)$ in terms of the beta distribution (40) and $F(s)$ in terms of the gamma distribution (39). For both of these the rejection algorithm is the fastest and this is what is implemented in almost all statistical numerical packages.

Lacking a covering distribution, one can always proceed by numerical inversion. Namely, one samples from the univariate uniform distribution, calling the deviate r. If we are using algorithms 1 or 2, we invert the cdf by finding t such that

$$I_N(t) = r. \tag{43}$$

For algorithm 3, we invert the algorithm so that

$$I(\tau) = r. \tag{44}$$

In the second case, we can work in terms of the distribution $G(s)$ (*i.e.* the cdf corresponding to the pdf $F(s)$) by finding

$$G(s) = r \tag{45}$$

and then using $\tau = 1/s$ as is evident from (34). We also remark that since r is uniformally distributed, we can equally well substitute I for I^* (or G for G^* in the three expressions above.) In order to invert the expressions above we can proceed by Newton iteration since the derivatives are given by the respective pdf functions. This can be efficient since we can determine the cdf and pdf at the same time for a given choice of variable and much of the computational effort is common and need not be duplicated. As an example, if we are using (24) and (25) to determine the pdf and cdf respectively, then we need only determine the function value $F(s)$ and the product st once. Similar logic also applies to the Cauchy representation.

To make the Newton method work efficiently, one also requires a scheme to estimate a good first guess for a given choice of r (as well as checks to make sure that we are converging to a solution and if not to proceed with some number of bisections.) Again, the best guess is always adapted to the specific copula at hand. Understanding the

asymptotic behaviour of the distributions in the limits of large and small arguments often allows for reasonable first guesses which can be effectively extended over the entire range of r values. For the Clayton copula the algorithms 1 and 3 are cleary superior to using any numerical inversion of the cdf, so it is not worth investing effort in the analysis of best first guesses. For the Gumbel copula, we have determined a number of properties of the function $F(s)$, including efficiently computed representations and asymptotic limits in both argument s and parameter α. We do not have space here to present all of these results and plan to present these results in a forthcoming publication.

For all algorithms, we verified that the procedure worked by testing that the marginal distributions generated from our samples of x are indeed uniform. We also tested dependence among the deviates by use of Kendall's tau statistic [Kendall and Stuart, 1979]. For the Clayton and Gumbel copulas we have

$$\begin{aligned} \tau &= \frac{\alpha}{\alpha+2} \qquad \text{Clayton} \\ &= \frac{\alpha-1}{\alpha} \qquad \text{Gumbel} \end{aligned} \tag{46}$$

This is a superior statistic to the standard correlation since it is zero if and only if the deviates are independent. For the exchangeable copulas, every choice of pairs of indices should have the same value of tau, within statistical error. Parenthetically we remark that from this statistic we see that the two copula families span the range from complete independence ($\tau = 0$) to complete dependence ($\tau = 1$).

We now present some timing statistics, namely the amount of time taken per sample using the algorithms presented above on a standard Pentium II 450MHz processor. We first present timing results for the Clayton copula in Table 1. We observe that algorithms 1 and 3 are virtually identical in terms of performance while algorithm 2 is three orders of magnitude slower. This is both because it uses inversion of the cdf, which is slow, and because of the integral representation of the distribution, which is also slow.

Clayton

Algorithm	# of Samples	Total cpu-time	cpu-time per sample
1	1,000,000	14.8s	1.48×10^{-5}s
2	1,000	15.0s	1.50×10^{-2}s
3	1,000,000	13.5s	1.58×10^{-5}s

Table 1. Timing results (in seconds) for the Clayton copula with $\alpha = 2$ and $N = 3$.

We also want to explore how well these algorithms work when we go to very large dimension. Below we present the timing results for a copula of dimension 100, which would be typical of the number of names in a CDO. The results are presented in Table 2 where we see that the algorithms 1 and 3 are essentially identical in terms of performance. There is about a factor of 10 loss of speed due to the necessity of making so many Poisson draws for each sample, but that is tolerable. By contrast algorithm 2 is about forty times slower than with $N = 3$ which we ascribe to the expressions (21) and

(23) becoming harder to work with as N becomes large, since the cost of the Poisson draws is negligible relative to the cost of the draw from $I_N(t)$. This points against its use for still larger N.

Clayton			
Algorithm	# of Samples	Total cpu-time	cpu-time per sample
1	1,000,000	208.2s	20.8×10^{-5}s
2	1,000	665.4s	66.5×10^{-2}s
3	1,000,000	207.9s	20.8×10^{-5}s

Table 2. Timing results (in seconds) for the Clayton copula with $\alpha = 2$ and $N = 100$.

We anticipate no problems going to arbitrarily high dimension with either algorithm 1 or 3 except for a linearly increasing time per draw. Since method 1 requires special properties while method 3 is generally applicable for any copula in which we can numerically determine $F(s)$, we would lean towards implementing 3 simply due to its wider applicability. Additionally it has the conceptual advantage that to obtain sampled values of τ, we sample from $F(s)$ without any regard to the copula dimension N. The information about N is simply encoded in the number of draws we make from a univariate Poisson distribution. By contrast the first two algorithms encode N directly in the distribution $I_N(t)$. This means that we need to explicitly worry about what happens as N gets large for example, a consideration not necessary for algorithm 3, for which robust scaling with dimension is manifest.

We next present analogous results for the Gumbel copula in Table 3. In this case we do not have recourse to algorithm 1. We note that algorithm 3 uses a numerical inversion of the numerically determined distribution $F(s)$ and is only a factor of four slower than the Clayton copula which is sampled using the rejection algorithm. As mentioned we plan to publish our results on the Gumbel copula and hope to explore whether an even more efficient scheme using rejection can be determined. In the meantime, this is certainly competitive with the Clayton copula. We note that algorithm 2 is slightly slower than for the Clayton but is similar.

Gumbel			
Algorithm	# of Samples	Total cpu-time	cpu-time per sample
1	N/A	N/A	N/A
2	1,000	22.09s	2.21×10^{-2}s
3	1,000,000	74.67s	7.47×10^{-5}s

Table 3. Timing results (in seconds) for the Gumbel copula with $\alpha = 2$ and $N = 3$.

Algorithm 2 is the only possibility when we do not know $F(s)$. However as we have discovered with the Gumbel copula, even not having $F(s)$ in closed form, it can

still be better to work with a numerical representation of it while using algorithm 3 than the much slower algorithm 2. Determination of a suitable representation of $F(s)$ obviously requires a specific analysis of the copula in question but at least in the case of the Gumbel copula that was clearly worth the effort. When we tried to increase the dimensionality, we found that algorithm 2 had problems beyond about 25. While this could probably be amended by optimising the integration contour for large N, it does underline the fact that this algorithm does not scale robustly. By contrast with algorithm 3, we found a million samples with $N = 100$ in 277.08s, which again compares favourably with the Clayton copula. (The fact that the ratio of times is much closer to unity than for $N = 3$ is evidence that most of the time is spent drawing from the Poisson distribution and not drawing from the $F(s)$ distribution.)

As a final word then, it would appear that algorithm 3 is the best in terms of conceptual simplicity and speed. It cleanly separates out the sampling of τ from any concerns about the dimensionality of the copula, meaning that it scales well with dimension. It is probably always worth finding an appropriate representation of $F(s)$ from the inverse Laplace transform of $f(x)$, even if it is numerical and not in a closed form. When this is not possible, algorithm 2 does work but is very slow and may fail for large dimension. Algorithm 1 may be competitive with 3 where it is available but it really only applies to special cases for which (25) can be integrated in closed form.

4. Fully Nested Copula

One generalisation of the multivariate Archimedean copula is given in [Joe, 1997] and also discussed in [Embrechts et al., 2001a]. It is given by $N-1$ distinct generating functions as

$$C(x_N, \cdots, x_1) = \phi_N^{-1}\left(\phi_N(x_N) + \phi_N\left(\phi_{N-1}^{-1}(\phi_{N-1}(x_{N-1}) + \cdots \right.\right. \\ \left.\left. + \ \phi_2^{-1}\left(\phi_2(x_2) + \phi_2(x_1)\right)\cdots\right)\right). \tag{47}$$

The structure is simple, if awkward to express in equations. We first couple x_1 and x_2. We then couple the copula of x_1 and x_2 with x_3. We then couple that copula with x_4 and so on. We shall refer to this situation as fully nested. In addition to the fact that alternate derivatives of each function $\phi_i^{-1}(x)$ need to alternate in sign (as in the exchangeable case) we have further constraints on the functions $\phi_{i+1}^{-1} \circ \phi_i(x)$, which we describe below. Also note that the $N-1$ copulas will be given by $N-1$ parameters. However, in general there are $N(N-1)/2$ pairings of variables and so there is still not enough structure to model all possible mutual dependencies amongst the variates. Nevertheless (47) is more general than (5).

In order to draw from this we could proceed by way of taking repeated conditional draws. The probability function for x_1 is unity so this is a trivial draw. We then have

$$\begin{aligned} I(x_2|x_1) &= \frac{\partial C_2(x_2,x_1)}{\partial x_1} \\ &\vdots \\ I(x_N|x_{N-1},\cdots,x_1) &= \frac{\partial^{N-1} C_N(x_N,\cdots,x_1)}{\partial x_{N-1}\cdots\partial x_1} \end{aligned} \tag{48}$$

The problem with this approach is that the partial derivatives get very complicated very quickly. As an example, for a 3-copula we find

$$\begin{aligned} \frac{\partial^2 C_3(x_3,x_2,x_1)}{\partial x_1 \partial x_2} &= f_3''(\phi_3)\phi_3'^2(f_2)f_2'^2(\phi_2)\phi_2'(x_2)\phi_2'(x_1) \\ &+ f_3'(\phi_3)\phi_3''(f_2)f_2'^2(\phi_2)\phi_2'(x_2)\phi_2'(x_1) \\ &+ f_3'(\phi_3)\phi_3'(f_2)f_2''(\phi_2)\phi_2'(x_2)\phi_2'(x_1) \end{aligned} \tag{49}$$

where we have defined $f_i(z) = \phi_i^{-1}(z)$ and for notational compactness we denote $\phi_2 = \phi_2(x_2) + \phi_2(x_1)$, $f_2 = f_2(\phi_2)$ and $\phi_3 = \phi_3(x_3) + \phi_3(f_2)$ as function arguments. For a Gumbel generating function $\phi_3(z) = (-\log z)^\alpha$. As an example of a higher derivative, we would find

$$\phi_3''(z) = \frac{\alpha(-\log z)^{\alpha-2}}{z^2}\left(\alpha - 1 - \log z\right). \tag{50}$$

There are expressions of similar complexity for $\phi_3'^2(z)$, $f_3''(z)$ and so on. Putting together (49) and (50), it is not hard to see that this approach will get unwieldy for dimensions beyond two or three.

Instead, inspired by the exchangeable case (7), we first define

$$\xi_i = \phi_i(x_i), \tag{51}$$

where for notational consistency we can define $\phi_1(z) = \phi_2(z)$. These are the fundamental variables we will work with and sample from. The probability distribution of the ξ_i is given by

$$P(\xi_N,\cdots,\xi_1) = (-1)^N \frac{\partial^N C_N(x_N,\cdots,x_1)}{\partial x_N \cdots \partial x_1}. \tag{52}$$

As for (8), we have a factor of -1 per dimension.

In terms of the ξ_i, we find

$$I_N^*(\xi_N,\cdots,\xi_1) = f_N(\xi_N + g_N(\xi_{N-1} + g_{N-1}(\xi_{N-2} + \cdots + g_2(\xi_2 + \xi_1)\cdots))), \tag{53}$$

where we have further introduced the "coupling functions"

$$g_N(z) = \phi_N \circ f_{N-1}(z). \tag{54}$$

The exchangeable copula can be understood as a special case in which all coupling functions are the identity function $g_i(x) = x$. C_N is a copula only if all the coupling functions g_i belong to the space of functions $\mathcal{L}_\infty^*$ where [Joe, 1997]

$$\begin{aligned} \mathcal{L}_\infty^* = \{&\omega : [0,\infty) \to [0,\infty) | \omega(0) = 0, \omega(\infty) = \infty, \\ &(-1)^j \omega^{(j)} \geq 0, j = 1,\cdots,\infty\}, \end{aligned} \tag{55}$$

effectively meaning that they map the positive real axis to the positive real axis and have alternate derivatives which alternate in sign. The properties of mapping the real axis to the real axis, mapping zero to zero and mapping infinity to infinity follow from the definition of the generating functions. However, the constraint of having alternatingly signed derivatives is nontrivial and the generating functions must be chosen carefully so as to satisfy this property. If it is violated, the copula will produce negative density functions. Embrechts et al. discuss this for the Gumbel copula [Embrechts et al., 2001a] and show that this additional constraint is satisfied if the successive generating functions $\phi_i(x)$ have parameters α_i which satisfy $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_N$, meaning the degree of dependence is greatest for the most deeply nested variates in (47).

We now set out to determine the distribution function of ξ_N, given all of the other variables. In order to do so, we introduce one more change of variables. We invoke the trivial map from ξ_i to ξ_i for all i but $N-1$; to replace ξ_{N-1} we define

$$t = \xi_{N-1} + g_{N-1}(\xi_{N-2} + \cdots + g_2(\xi_2 + \xi_1)\cdots). \tag{56}$$

The merit of this change of variables is that I_N^* only depends on ξ_N and t so that the derivatives become simpler to compute.

$$\begin{aligned}
\frac{\partial I_N^*}{\partial \xi_N} &= \frac{\partial I_N^*}{\partial \xi_N} \\
\frac{\partial I_N^*}{\partial \xi_{N-1}} &= \frac{\partial I_N^*}{\partial t} \\
\frac{\partial I_N^*}{\partial \xi_{N-2}} &= \frac{\partial t}{\partial \xi_{N-2}} \frac{\partial I_N^*}{\partial t} \\
&\vdots \\
\frac{\partial I_N^*}{\partial \xi_1} &= \frac{\partial t}{\partial \xi_1} \frac{\partial I_N^*}{\partial t}.
\end{aligned} \tag{57}$$

The various partial derivatives $h_i(\xi) \equiv \frac{\partial t}{\partial \xi_i}$ have the useful property that they only depend on ξ_1 to ξ_{N-2} and on neither t nor ξ_N. Therefore successive differentiations with respect to t do not affect them. We can then write

$$P(\xi_N, \cdots, \xi_1) = (-1)^N \left(\prod_i h_i(\xi) \right) \frac{\partial}{\partial \xi_N} \frac{\partial^{N-1}}{\partial t^{N-1}} f_N\left(\xi_N + g_N(t)\right). \tag{58}$$

First, we integrate over $\int_0^\infty d\xi_N$, which is trivial since (58) is a perfect derivative in terms of ξ_N so that

$$P(\xi_{N-1}, \cdots, \xi_1) = (-1)^{N-1} \left(\prod_i h_i(\xi) \right) \frac{\partial^{N-1}}{\partial t^{N-1}} f_N\left(g_N(t)\right). \tag{59}$$

The ratio of the previous two expressions is the conditional density

$$P(\xi_N | \xi_{N-1}, \cdots, \xi_1) = -\frac{\frac{\partial}{\partial \xi_N} \frac{\partial^{N-1}}{\partial t^{N-1}} f_N\left(\xi_N + g_N(t)\right)}{\frac{\partial^{N-1}}{\partial t^{N-1}} f_N\left(g_N(t)\right)}. \tag{60}$$

We can then integrate this expression with respect to ξ_N to determine the conditional cdf. Again, this is a perfect derivative and is trivially integrated:

$$I^*(\xi_N|\xi_{N-1},\cdots,\xi_1) = \frac{\frac{\partial^{N-1}}{\partial t^{N-1}} f_N\left(\xi_N + g_N(t)\right)}{\frac{\partial^{N-1}}{\partial t^{N-1}} f_N\left(g_N(t)\right)}. \tag{61}$$

This is manifestly normalised since it equals unity for $\xi_N = 0$.

Equation (61) is the main result of this section. However it does merit some discussion. Firstly we can see that the right hand side is only a function of ξ_N and t. However, we can also interpret t in terms of the copula C_{N-1} as $t = \phi_{N-1}(C_{N-1})$. In other words, (61) gives the distribution of ξ_N conditional on the copula function of the previous draws, which is an appealing recursive property. Second, we can simplify the denominator through the relation $f_N(g_N(t)) = f_n \circ g_N(t) = f_{N-1}(t)$. In fact, applying this same identity to (59), it is not difficult to see that that equation is consistent with the variables ξ_i to ξ_{N-1} having the copula C_{N-1}, as they must.

Then the algorithm is that we first draw a value x_1 randomly from zero to unity. We then determine ξ_1 and from that determine a random draw ξ_2 using (61). We then calculate $x_2 = \phi_2(\xi_2)$ and the value $t = \phi_2(C_2(x_1, x_2))$. We can then make a conditional draw for ξ_3 conditional on the value of t, using (61). We continue until we have N drawn values.

As mentioned, an alternative algorithm for two dimensions is presented in [Embrechts et al., 2001a]. It is distinct from what we have derived here but does share the property that one first makes a draw for x_1 and then makes a draw for x_2 conditional on x_1. Note that in two dimensions there is no distinction between exchangeable and nested Archimedean copulas, so the need for distinct algorithms only arises in greater than two dimensions.

Unfortunately this algorithm still requires determining high order derivatives in (61). There is nevertheless an advantage over directly working with the distribution in terms of the original variables x. Namely we only need to concern ourselves with derivatives with respect to one variable, t, and not with respect to all of the $x_1, \cdots, x_{N-1}$. We believe that (61) represents the most parsimonious sampling algorithm and that the complication of having to determine the high order derivatives as a function of t is an unavoidably intrinsic property of this class of copulas. Possible methods to calculate the required derivatives include using a symbolic manipulation package with automatic code generation and using "automatic differentiation" to generate code for derivatives (for reviews see [Rall, 1981] and [Corliss et al., 2002].)

One can also use the Cauchy formula supplemented with asymptotic relations for the derivatives in the limits of large and small arguments. Specifically we can write

$$I^*(\xi_N|t) = \frac{(N-1)!}{2\pi i \frac{\partial^{N-1}}{\partial t^{N-1}} f_{N-1}(t)} \oint \mathrm{d}z \frac{f_N\left(\xi_N + g_N(z)\right)}{(z-t)^N} \tag{62}$$

together with a corresponding expression for $P(t)$, its derivative. We invoked the last method to confirm that this algorithm works for the Gumbel copula. In Table 4 we present timing statistics for a selection of dimensions. We observe that for two or

three dimensions it is competitive with algorithm 2 of the exchangeable copula case. The time is roughly linear in dimension up to dimensin 5. Beyond that we found it to be very slow, presumably due to the increasing complexity of (62) with dimension. With effort, this could be improved upon by better adapted integration contours, for example. Nevertheless this is indicative that this method probably would be difficult to scale beyond ten or twenty dimensions. An interesting direction to study, if one wanted to use these nested copulas for large dimension, is whether one can find appropriate covering distributions parameterised by N and t and the relevant copula parameters in order to make use of the rejection algorithm.

Gumbel			
Dimension	# of Samples	Total cpu-time	cpu-time per sample
2	1,000	10.5s	1.05×10^{-2}s
3	1,000	27.7s	2.77×10^{-2}s
4	1,000	41.9s	4.19×10^{-2}s
5	1,000	58.1s	5.81×10^{-2}s

Table 4. Timing results (in seconds) for the fully nested algorithm using the Gumbel copula with $\alpha = 2$ and $N = 3$.

5. Partially Nested Copula

There are alternate multivariate extensions to the fully nested copula of the previous section. They can be understood as composites of the exchangeable copula and the fully nested copula. The lowest dimension in which there is a distinct copula of this class is four, for which the copula function is

$$C(x_4, x_3, x_2, x_1) = \phi^{-1}\left(\phi\left(\phi_{12}^{-1}(\phi_{12}(x_1) + \phi_{12}(x_2))\right)\right. \\ \left. + \phi\left(\phi_{34}^{-1}\left(\phi_{34}(x_3) + \phi_{34}(x_4)\right)\right)\right). \tag{63}$$

Again, the equation looks complicated although the logic is straightforward. We first couple the two pairs x_1, x_2 and x_3, x_4 with distinct copulas generated by ϕ_{12} and ϕ_{34} respectively. We then couple the two copula functions using a third generating function ϕ. Joe discusses how this structure and that of the fully nested copula can be understood from distinct multivariate Laplace transforms [Joe, 1997]. This distribution is exchangeable between x_1 and x_2 and also between x_3 and x_4 and for that reason can be understood as intermediate between the fully exchangeable copula and the fully nested copula. Like (47) for $N = 4$, (63) is generated by three distinct generating functions.

It is not difficult to imagine other patterns of nesting in higher dimensions. It can be notationally overwhelming to attempt to express symbolically the most general case. Nevertheless in the following section we do present one particular choice of nesting to reflect hierarchical structure among the random deviates. Rather than writing an algorithm for the most general case of partial nesting, we shall focus on the particular

multivariate distribution (49). The pattern of how to handle more complicated nestings should then be clear.

We first define $\xi_i = \phi_{12}(x_i)$ for $i = 1, 2$ and $\xi_i = \phi_{34}(x_i)$ for $i = 3, 4$ and then $t = \xi_1 + \xi_2$ and $s = \xi_3 + \xi_4$ as well as u_t and u_s which span the constant t and s surfaces respectively. Appropriately enough, our approach for this problem is a composite of the approaches for the exchangeable copula and the fully nested copula. As for the exchangeable copula, we seek a strategy to make random draws of the variables s and t from which we can then trivially determine random deviates of ξ_i and hence of x_i. As for the fully nested copula, we do this using conditional arguments; we first draw a value of of t from the two dimensional copula of x_1 and x_2 and then draw a value of s conditional on it. We now present the details.

First we define the functions $f(x) = \phi^{-1}(x)$ (and similarly for $f_{12}(x)$ and $f_{34}(x)$) and coupling functions

$$\begin{aligned} g_{12}(x) &= \phi \circ f_{12}(x) \\ g_{34}(x) &= \phi \circ f_{34}(x) \end{aligned} \tag{64}$$

in terms of which

$$C(x_4, x_3, x_2, x_1) = f\left(g_{12}(t) + g_{23}(s)\right). \tag{65}$$

There are similar constraints on the coupling functions as in the fully nested copula [Joe, 1997].

The joint pdf of the ξ variables is the fourth derivative of the copula function with respect to the ξ variables (as for both the exchangeable and fully nested copulas.) Under the change of variables to t, s, u_t and u_s, this becomes

$$P(t, s, u_t, u_s) = \frac{\partial^2}{\partial t^2}\frac{\partial^2}{\partial s^2} f\left(g_{12}(t) + g_{23}(s)\right). \tag{66}$$

As for the exchangeable copula, even though this does not depend on the u variables, we still get non-trivial factors of t and s when integrating over their measures to determine the marginal distribution of just t and s. That is

$$P(t, s) = ts\frac{\partial^2}{\partial t^2}\frac{\partial^2}{\partial s^2} f\left(g_{12}(t) + g_{23}(s)\right). \tag{67}$$

From the discussion in the exchangeable section, we know that the marginal distribution for t alone is $P(t) = t\frac{\partial^2}{\partial t^2} f_{12}(t)$. Therefore the conditional density is given as the ratio which can then be integrated with respect to s to determine the complement of $I(s|t)$ as

$$I^*(s|t) = \frac{1}{f''_{12}(t)}\left(\frac{\partial^2}{\partial t^2} f\left(g_{12}(t) + g_{23}(s)\right) - s\frac{\partial^3}{\partial s \partial t^2} f\left(g_{12}(t) + g_{23}(s)\right)\right). \tag{68}$$

Using (64), we observe that this equals unity for $s = 0$ and hence is normalised. In practise, we can expand out all of the derivatives using the chain and product rules or evaluate the derivatives in some numerical manner.

The algorithm is then to draw a value of t as in the exchangeable section. Next draw a value s conditional on the drawn value of t using the conditional cdf above. Next find

values of ξ_i consistent with the sums equalling t and s (as for the exchangeable copula discussed in subsection 3.3.) Finally map the values ξ_i to the corresponding x_i.

6. Hierarchical Copulas

We end this paper on a more speculative note. The different nesting combinations discussed in the previous two sections suggest that a hierarchical structure among the random variates can be naturally implemented in the copula function itself. To make this explicit we use an example from finance but the same ideas apply to any application with a hierarchical structure.

It is common to categorise obligors by their credit rating and industrial sector. For example there is a growing market in tranches of collateralised loan obligations in which different counterparties assume (for a fee) the credit risk associated with default of some fraction of a large basket of names. The payoff function for this derivative is strongly dependent on the assumed dependency structure of the large basket. Rather than treating every obligor in a name-specific way we can categorise them by credit rating and industrial sector. Then we can treat all obligors with the same rating and sector using an exchangeable copula function. This can be expressed symbolically as

$$C_{ij} = \phi_{ij}^{-1}\left(\sum_k \phi_{ij}(x_{ijk})\right) \tag{69}$$

where i labels the sector, j labels the rating and k is a dummy index labelling the obligors within a given rating/sector combination. We have defined a rating/sector specific generator $\phi_{ij}(x)$.

We would then want to combine across sectors for a given rating (alternately we could combine across ratings for a given sector, the logic is the same and this amounts to a modelling choice.) We can do this with a new set of generators $\psi_i(x)$, one per rating, so that the copula for all obligors with a given rating would be

$$C_i = \psi_i^{-1}\left(\sum_j \psi_i(C_{ij})\right). \tag{70}$$

The dependence on the variates x is implicitly contained in the copula functions C_{ij}.

The copula for the entire basket is then obtained by combining across all ratings. We assume that the generating function for this is $\chi(x)$ so that we have

$$C = \chi^{-1}\left(\sum_i \chi(C_i)\right). \tag{71}$$

This choice of multivariate copula has both nesting and exchangeable elements. There are constraints on the various generating functions analogous to (55) but presumably they satisfy the reasonable property of requiring less dependence as we consider higher orders of nesting, as we remarked upon for the fully nested Gumbel copula. Obviously these questions need study but this would go beyond the scope of the present paper. Whether this hierarchical copula has enough structure to accurately model the behaviour

of a basket is an interesting question. For example, it assumes that within a rating/sector combination all obligors are identically correlated (more generally are exchangeable.) It also assumes that across different sectors and ratings the dependencies are the same; all industrial obligors are correlated the same with all financial obligors and with all energy obligors, for example. The ideas of this paper can be readily generalised to allow for sampling from these distributions.

This structure generalises in an obvious manner. For example if we wished to add a new attribute, such as the country where the obligor is based, we just add one more index to the obligors and introduce one more level of hierarchy in the nesting of the copulas. We can also impose a hierarchical structure directly on the correlation matrix of a Gaussian copula and it would be interesting to compare the relative performance of the two approaches. In a related vein, an alternative method for lessening the number of parameters and effective dimensionality is to use a factor copula [Laurent and Gregory, 2002] and it would be interesting to explore the degree of overlap between that approach and the one outlined here. We plan to explore these ideas more fully in a later publication.

7. Conclusion

In this paper we have introduced three algorithms for sampling from exchangeable Archimedean copulas. In the course of doing this we have found several integral representations for the distribution of the copula function as well as a generating function expansion. We also derived an expression for the copula distribution function in the infinite dimensional limit which directly relates it to the distribution whose Laplace transform yields the generating function. We hope to expand on this in a subsequent paper, including developing various asymptotic arguments for large and small arguments as well as on large but not infinite dimension. Already we see the utility of the infinite dimensional limit in providing a particularly elegant sampling algorithm in 3.4. This was the algorithm we decided upon as the best technique due to its separation of the sampling of τ from any consideration about the dimensionality of the problem.

We also provided analysis and algorithms for sampling from inexchangeable Archimedean copulas by working in terms of new variables. These algorithms still require direct determination of high order derivatives of the generating functions, but only in terms of one variable. Therefore they are simpler than direct use of consecutive conditional draws of the original x_i variables and make use of what we believe is the simplest possible structure for purposes of sampling.

We have also suggested a class of Archimedean copulas which reflects the hierarchical structure among the random variates and discussed a possible application of this idea in the area of credit derivatives.

Appendix A. Numerical determination of contour integrals

We have expressed a number of our results in terms of contour integrals. Despite the fact that determination of contour integrals is a relatively straight-forward numerical task, we are not aware of its being extensively discussed in standard texts on numerical methods; a discussion of a specific application to a non-closed contour can be found in [Press et al., 1992]. Therefore for completeness, we briefly discuss how to do this. Imagine we wish to perform the following contour integral

$$I = \oint_C \mathrm{d}z f(z) \tag{A.1}$$

over some closed contour C. The first point is that we can always parameterise the contour as $z(t)$ which is a complex function of a real parameter t. Without loss of generality, t can be defined between 0 and 1. We then find

$$\begin{aligned} I &= \int_0^1 \mathrm{d}t \frac{\mathrm{d}z}{\mathrm{d}t} f(z(t)) \\ &= \int_0^1 \mathrm{d}t F(t), \end{aligned} \tag{A.2}$$

where we have defined the function $F(t) = \frac{\mathrm{d}z}{\mathrm{d}t} f(z(t))$. We can then apply standard methods of numerical integration with the proviso that the integrand is complex, but this is a relatively minor complexity.

It is worth noting however that assuming the contour crosses no singularities (which is almost certain to be true in any practical application) and is an analytic function of t, then $F(t)$ is an analytic and periodic function of t. In that event the best way to integrate over t is to approximate it as a discrete sum at equally spaced points t which by an extension of the Euler-Maclaurin summation formula, converges exponentially with the number of sampled points. In particular

$$I_N \equiv \frac{1}{N} \sum_{i=0}^{N-1} F\left(\frac{i}{N}\right) \tag{A.3}$$

approaches I faster than any reciprocal power of N as $N \to \infty$ [Hirayama, 2001]. This is particularly nice approach due to the recursion formula

$$I_{2N} = \frac{1}{2}\left(I_N + \frac{1}{N} \sum_{i=0}^{N-1} F\left(\frac{i+1/2}{N}\right)\right). \tag{A.4}$$

(This just interleaves the set of points used in determining I_N with a new set spaced midway between.) We can start at some small reasonable value of N, such as 4, and progressively double N. We stop when the sum has converged to within our desired tolerance. In practise a relatively small handful of function evaluations should be sufficient to determine the integral to typical desired accuracies. This approach can be further supplemented with "Richardson's deferred approach to the limit" where at each choice of $1/N$, we use all previously determined values to extrapolate the result to

$1/N \to 0$, which is the idea behind Romberg integration and the Bulirsch-Stoer method for integrating ordinary differential equations [Press et al., 1992].

What remains is to choose the contour which can be integrated the most efficiently and to choose the parameterisation $z(t)$. There are no clearcut choices and the best choice is presumably specific to the problem at hand. We therefore outline some general principles which can be considered. We want to avoid any kinks in the contour or else we lose the property that the integrand of (A.2) is analytic in t. It is then not to hard to see that even if analytic, to the extent that the contour has large curvature, the convergence will be slower. This suggests that a circular contour will typically be the best choice, if this is possible given the topology of the contour. Convergence will typically be fastest if the $F(t)$ has as little structure as possible meaning no sharp changes in amplitude over narrow ranges of t. For this reason, we will also typically want that any pole be at the centre of the contour, assuming there is just one pole. If there are multiple poles we choose a point somewhere in the middle of them.

We must then choose the radius of the contour. If we make it too small, we run the risk of potentially having large magnitudes in the sum (A.3) thereby requiring precise cancellation amongst the various terms, leading to a possible loss of numerical precision. On the other hand, we do not want to allow it to be so large that the contour approaches other singularities in the complex plane. This is best explored numerically as there is little beyond this that can be said of any generality. One reasurring result is that the method is very robust assuming one does keep the contour away from any singularities and not let it get too large. As an example, for the Cauchy integral (21), a good choice is to centre the contour at $z = t$ and to select a radius of $Nt/(N+1)$. That choice of radius assures us that the integrand is close to its minimum value when the contour crosses the real axis for $z < t$. This minimises the possibility that there are large fluctuations in magnitude as we traverse the contour, thereby minimising the computational effort.

Once we have selected the contour, we are left with the choice of determining $z(t)$. Assuming that it is a circular contour, in general a robust choice will be that the angular position in the complex plane is linear in t so that $z(t) = z_0 + R\exp(2\pi it)$, where z_0 is the centre of the contour and R is the radius; this is the parameterisation we have used in the work in this paper. However even for a given contour other choices of parameterisation could possibly be more efficient; this is a detailed question which would be specific to the problem at hand.

Integrals which are not closed but extend to infinity, such as arise from inverse Laplace transforms, can be mapped to either a finite or an infinite integral over a real parameter t. This is then a standard problem in integration. Questions about the choice of contour and of the parameterisation $z(t)$ still remain. For example, it is often best to select a contour for which the integrand, or some chosen part of the integrand, has constant phase.

Acknowledgments

I wish to thank Jeff Boland, Tom Hurd, Harry Joe and Shaloub Razak for many useful comments and discussions.

References

M. Abramowitz and I. Stegun, editors. *Handbook of Mathematical Functions*. Dover Publications, New York, 1965.

J. Ahrens and U. Dieter. Computer methods for sampling from gamma, beta, poisson and binomial distributions. *Computing*, 12:223–246, 1974.

P. Barbe, C. Genest, K. Ghoudi, and B. Rémillard. On Kendall's process. *Journal of Multivariate Analysis*, 58:197–229, 1996.

E. Bouyé, V. Durrleman, A. Nikeghbali, G. Riboulet, and T. Roncalli. Copulas for finance: A reading guide and some applications. Preprint of the Groupe de Recherche Opérationnelle, Crédit Lyonnais, 2000.

G. Corliss, C. Faure, A. Griewank, and L. Hascoet, editors. *Automatic Differentiation of Algorithms*, 2002. Springer-Verlag.

L. Devroye. *Non-uniform Random Variate Generation*. Springer-Verlag, New York, 1986.

P. Embrechts, F. Lindskog, and A. McNeil. Modelling dependence with copulas and applications to risk management. Preprint of the ETH, September 2001a.

P. Embrechts, A. McNeil, and D. Straumann. Correlation and dependence in risk management: Properties and pitfalls. In M. Dempster and H. Moffat, editors, *Risk Management: Value at Risk and Beyond*. Cambridge University Press, Cambridge, 2001b.

R. Frey and A. J. McNeil. Dependent defaults in models of portfolio credit risk. in preparation, May 2003.

C. Genest and L. Rivest. Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88:1034–1043, 1993.

H. Hirayama. Numerical integration on the complex plane. *IPSJ SIGNotes Numerical Analysis Abstract*, No.022 - 003, 2001.

W. Höffding. Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Seminars und des Instituts für Angewandte Mathematik der Universität Berlin*, 5: 181–233, 1940.

H. Joe. *Multivariate Models and Dependence Concepts*. Number 37 in Monographs on Statistics and Applied Probability. Chapman and Hall, London, 1997.

M. Kendall and A. Stuart. *Handbook of Statistics*. Griffin and Company, London, 1979.

C. Kimberling. A probabilistic interpretation of complete monotonicity. *Aequationes Mathematicæ*, 10:152–164, 1974.

J.-P. Laurent and J. Gregory. Basket default swaps, cdo's and factor copulas. Available at http://laurent.jeanpaul.free.fr, 2002.

D. X. Li. On default correlation: a copula function approach. *Journal of Fixed Income*, March, 2000:43–54, March 2000.

A. W. Marshall and I. Olkin. Families of multivariate distributions. *Journal of the American Statistical Association*, 83:834–841, 1988.

R. Nelsen. *An Introduction to Copulas*. Number 139 in Lecture Notes in Statistics. Springer, New York, 1999.

F. Oberhettinger. *Tables of Fourier Transforms and Fourier Transforms of Distributions*. Springer-Verlag, Berlin, 1990.

W. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, 2 edition, 1992.

L. B. Rall. *Automatic Differentiation: Techniques and Applications*, volume 120 of *Lecture Notes in Computer Science*. Springer Verlag, Berlin, 1981.

S. Razak. Default intensities from copulas. in preparation, 2003.

I. J. Schoenberg. Metric spaces and completely monotone functions. *Ann. Math.*, 39: 811–841, 1938.

P. J. Schönbucher and D. Schubert. Copula-dependent default risk in intensity models. Preprint of Department of Statistics, Bonn University, April 2001.

A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris*, 8:229–231, 1959.